**Challenge Set 1 SOLUTIONS.**

**Problem 1 (Merit)**

On a particular stretch of road, accidents occur at an average rate of one per 4 working days.

1. Poisson distribution, assuming that:

* Accidents occurring at random.

This assumption is questionable because will probably be *predictable* seasonal patterns (seasonality) with accidents occurring more frequently during certain times of the day (e.g. “rush hour”) and certain months of the year (e.g. the colder months when more people drive).

* Accidents occurring independently of one another is another questionable assumption (e.g. hazardous driving conditions on a particular day would violate this assumption as this would increase the likelihood of multiple accidents, linked by common risk factors).
* The probability of an accident occurring is proportional to the length of time of observation (probably OK as long as each interval of time measured includes 24 hours, i.e. containing all levels of activity throughout a day & night).
* That the probability of two or more accidents occurring simultaneously is negligible (probably, as long as we classify related crashes such as a pile-up as a *single* accident).

1. **i**. Poisson with ****= ¼ accidents per day on average.

P(no accident on a given day) = e****= e****0.25 = 0.7788007831

P(no accident on Monday and Tuesday) = P(no accident on a given day)2

= 0.77880078312

P(at least one accident on a given day) = 1 – P(no accident on a given day)

= 1 – e****0.25

= 0.2211992169

P(at least one accident on each of Wed, Thurs, and Fri) = P(at least one accident on a day)3

= 0.22119921693

**P(no accident on Mon, Tues, and**

**at least one on each of Wed, Thurs, Fri) =** 0.77880078312 × 0.22119921693

**= 0.006565 (4sf)**

**ii. Probability that at least 3 days in the 5 are accident free.**

This is **Binomial** since there is a fixed number of independent trials: 5 days, and each day has 2 possible outcomes: accident free or not.

**Parameters:** ***n*** = 5 days. ****** = P(no accident on a given day) = 0.7788007831

Let *X*= number of days that are accident free.

P(*X* > 3) = 1 – P(Binomial: *X* < 2)

= 1 – 0.075497 (*G.C.*)

= **0.924503**

**Problem 2 (Merit and Excellence)**

A large airline suffers from randomly-occurring equipment faults requiring replacement aircraft on 90% of all days.

Flights need to be cancelled if there are three or more cases of faulty aircraft per day, because there are only two replacement aircraft.

1. **Poisson** – **discrete** events occurring at random within a **continuous interval**.

Asked the value of ****, the mean number of faults per day. So an inverse problem.

**Let *X*: Number of faults in a given day.**

P(*X* > 1) = 0.9 (from info in question) So P(*X* = 0) = 1 – 0.9

= 0.1

So, using the Poisson formula: P(*X*=*x*) = and subbing in *x*=0.

= 0.1. **** e**** = 0.1

****loge(e**** ) = loge0.1

****logee = loge0.1

** ** = loge0.1 *since logee = 1*

** ** = –loge0.1

** ** **= 2.302585093 faults per day on average.**

**(b)**Given that at least one replacement aircraft had to be used, calculate the *conditional* probability that only two flights had to be cancelled

*2 replacement aircraft, so 1 flight cancelled if there are 3 faults (both spares used). So 2 flights would be cancelled if there were 4 faults in a day.*

P(2 flights cancelled │at least one replacement aircraft needed)

= P(*X* = 4 │ *X* > 1)

= 

= 

=  *(G.C.)*

= **0.1301** (4sf)

**(c)** Calculate the mean number of replacement aircraft used per day.

Let ***Y*** : Number of **replacement aircraft needed** in a given day. Then the probability distribution of *Y*  is given by:

|  |  |  |  |
| --- | --- | --- | --- |
| ***y*** | 0 | 1 | 2 *(Note: this is occurs if there are 2 or more faults)* |
| **P(*Y*=*y*)** | 0.1 | 0.23025 | 0.66975 |

Mean number of replacement aircraft = E(*Y*)

= 0(0.1) + 1(0.23025) + 2(0.66975)

= **1.56975 aircraft needed per day on average.**

**Problem 3 (Merit and Excellence)**

Faults occur randomly on a printing machine, which produces wallpaper which is later wound up into rolls. Any roll with a defect is rejected; this happens to 16% of the rolls.

1. Calculate the mean number of faults per roll.

**Poisson** – **discrete** events occurring at random within a **continuous interval**.

Asked the value of ****, the mean number of faults per roll of wallpaper. So an inverse problem.

**Let *X*: Number of faults on a given roll of wallpaper.**

P(*X* > 1) = 0.16 (from info in question)

So P(*X* = 0) = 1 – 0.16

= 0.84

So, using the Poisson formula: P(*X*=*x*) = and subbing in *x*=0.

= 0.84. **** e**** = 0.84

****loge(e**** ) = loge0.84

****logee = loge0.84 *but logee = 1*

** ** = loge0.84

** ** **= 0.1743533871 faults per roll of wallpaper on average.**

1. What is the probability that a **rejected** roll has **at least two faults**.

*Rejected if it has 1 or more faults.*

P(At least 2 faults │Rejected) = P(*X* > 2 │ *X* > 1)

= 

= 

= 

=  *(G.C.)*

= 

= **0.08469** (4sf)

1. To cut down the number of rejected rolls, the manufacturer decides to halve the amount of wallpaper on a roll. Find the percentage of rolls which are now rejected.

**** **= ** = **0.087176699357** **faults per roll of wallpaper on average.**

As a Poisson variable, the frequency of a fault is proportional to the size of the interval in which it occurs. So halving the amount of paper in the roll will halve the mean rate of faults.

Then P(*X* > 1) = 1 – P(*X* = 0)

= 1 – 0.916515139

= **0.08348 (4sf). So expect that 8.348% of rolls will now be rejected**.

**Problem 4 (2010 NCEA Merit & Excellence)**

1. A farmer grows onions and plants them in rows which are 50 metres long.

The farmer needs to know the number of onions that will be diseased.

To investigate this, he divides one row of onions in a paddock into 5 sections, each 10 metres long, and counts the number of diseased onions in each section.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **1st section** | **2nd section** | **3rd section** | **4th section** | **5th section** |
| Number of diseased onions | 13 | 17 | 9 | 9 | 12 |

Use the information in the table above to find the probability that, in a **5 metre section**, more than 4 onions will be diseased.

**Mean** number of diseased onions **per 10 metre section** = = 12

**** Assuming that the diseased onions occur independently and at random and that their frequency is proportional to the length of row, we can model it using the **Poisson Distribution**.

Then the mean number of diseased onions in a **5 metre section** is **** = =**6.**

Then if *X*: The number of diseased onions in a given 5 metre section,

P(*X* > 4) = 1 – P(*X <* 4)

= 1 – 0.28505 *(G.C.)*.

= **0.71495**.

1. The farmer estimates that, for each day of onion harvesting, the probability of two or more workers calling in sick is 0.2 and that the probability of one worker calling in sick is 0.35.

Find the expected number of workers calling in sick each day of onion harvesting.

This is an inverse problem, need to find **** the expected number (i.e. mean) who call in sick per day.

Let *X*: Number of workers calling in sick in a given day.

From info in question: P(*X*=1) = 0.35; P(*X* > 2) = 0.2

**** P(*X*=0) = 1 – 0.35 – 0.2

= 0.45. i.e. estimate that, on 45% of days, none of the workers calls in sick.

So, using the Poisson formula: P(*X*=*x*) = and subbing in *x*=0.

= 0.45. **** e**** = 0.45

**** = loge0.45 *since logee = 1*

** ** **= 0.7985 workers calling in sick per day on average** *(to 4sf)***.**

**Problem 5**

The final series of the America’s Cup between two yachts is organised so that the winning yacht is the first one to win five races out of a possible total of nine races.

Calculate the probability that **not all nine races** are needed to find the winner. Assume that each yacht is equally likely to win each race, and that the results of the races are independent.

P(Not all 9 races needed) = 1 – P(9 races needed)

= 1 – P(score is “*4-all”* after 8 races)

= 1 – P(Binomial dist: *X=4, n=8, =0.5*)

Where *n*=number of races (8), *X*=number of wins to each yacht (4),

*=*probability of each yacht individually winning a given race (0.5).

P(Not all 9 races needed) = 1 –

= 1 –

= 1 – 0.27343

= **0.7266 (4SF)**

**Problem 6**

Andy and Bernie are shooting arrows at a target. At each turn, the player with an arrow closest to the centre scores a point. Each player is equally likely to be closest.

They take 25 turns, and the player with the highest score wins.

After 10 turns Andy has a score of 7.

What is the probability that **Bernie** wins the game?

**A** stands for Andy; **B** stands for Bernie.

After 10 turns, score must be 7:3 to A.

Therefore B is down 4 points. So he must score at least 5 points more than A in the rest of the game.

There are 15 turns left (25-10). B must win at least 10 of them.

The probability of this occurring is Binomial:

P(B wins at least 10 of the remaining 15 turns) =

= **0.15087** (G.C.)

**Problem 7**

The emergency locator beacon on a ship uses batteries which are installed when needed from a packet of 11.

The packet is labelled ‘*average life = 12 hours*’.

The life of these batteries is normally distributed with standard deviation 0.7 hours.

1. Calculate the minimum length of time you would expect 95% of these batteries to last for.

Inverse Normal Distribution problem: With ****** = 12 hours, ****** = 0.7 hours.

Need to find *x* such that **P(*X* > x) = 0.95.**

i.e.  = 0.95. i.e.  = 0.45

Using inverse tables, we get *z* = 1.645 (meaning -1.645). . Solving gives ***x =* 10.8485**

**So expect 95% of batteries to last for at least 10.8 hours (i.e. at least 10 hours, 50 mins).**

1. An emergency arises, and all 11 of the batteries are used, one after the other. Calculate the probability that the *total* length of time that the 11 batteries last exceeds 135 hours.

Let ***T***: Total length of time that all 11 batteries last for.

***T*** has mean ***E(T)* = 11(12)** and **standard deviation** 

**= 132 hours** = 

= **2.321637353 hours**

**P(*T* > 135) = 0.098145** *(G.C.)*

1. Calculate the probability that exactly 4 of the batteries from the packet each lasts more than 13 hours.

**Probability that an individual battery lasts > 13 hours:**

Let *X* = time battery lasts. *X* is a **normally distributed** variable with ****** = 12 hours, ****** = 0.7 hours.

P(*X* > 13) = 0.076563 *(G.C.)*

**Number of batteries in a packet lasting > 13 hours:**

Let *Y*: Number of batteries from packet that last > 13 hours.

*Y* is a **Binomial** random variable (11 independent trials, each of which has 2 poss. outcomes).

Parameters are ***n*** = 11 batteries in packet. ****** = P(an individual battery lasts > 13 hours)

= 0.076563

P(*Y* = 4) = 11C4 (0.076563)4 (1 – 0.076563)7

= **0.0064929**. This is the probability that exactly 4 of the batteries last for > 13 hours.

**Problem 8 - The Geometric & Negative Binomial Distributions** *(Can feature in Scholarship)*

You throw a fair six-faced die, observe which number is facing upward, and then throw it again, repeating this until you roll a 6 for the first time:

1. Calculate the probability that it takes:

**i.** exactly 2 throws to get your first 6:  **ii.** exactly 3: 

1. Write down a **calculation** for the following probabilities but **do not evaluate them**:
   * 1. The probability that it takes exactly **5** throws to get your first six. 
     2. The probability that it takes exactly **10** throws to get your first six. 
2. Write down an expression for the probability that it takes exactly ***m*** throws to get your first six. 
3. The **Geometric Distribution** is another probability distribution that is related to the Binomial Distribution.

If independent trials are carried out, each of which has two possible outcomes defined as ‘success’ and ‘failure’ with constant probability ****** of success, we use the Geometric Distribution to model the number of trials ***m*** up to and including the first success.

Based on your answer to (c), write down an expression for **P(*M=m*)**, the probability that the ***mth*** trial results in the **first** **success**. P(*M* = *m*) = 

1. Returning to the dice experiment, what is the probability that:
   1. The third roll of the die results in the **third six** (i.e. all 3 are sixes)? 
   2. The **fourth** roll of the die results in the **third six?**

= P(Binomial: 2 successes in 3 trials) × P(success on 4th trial)

= 

= or 0.01157 (4sf)

* 1. The **seventh** roll of the die results in the **third six?** 0.03349 (4sf)
  2. The ***m***th roll of the die results in the **third six?** 
  3. The ***m***th roll of the die results in the ***k***th **six**? 

1. Write down a simplified expression for the probability that, in a sequence of independent trials with probability of success ****** in each, the ***m*th** trial results in the ***k*th** success (this is known as the ‘*Negative Binomial Distribution’*).

P(*M* = *m, K=k*) = P(Binomial: *m-1* successes in *k-1* trials) × P(success on *kth* trial)

P(*M* = *m, K=k*) = 

Which simplifies to P(*M* = *m, K=k*) = 

**Problem 9 (2004 NCEA Merit & Excellence)**









**Problem 10 *(A classic Scholarship problem)***

On average, 30% of smokers can give up successfully on the first attempt.

In a random sample of ***n*** smokers who intend to make their first attempt at giving up smoking, find ***n*** so that there is 0.9 probability that at least 30 of them give up successfully.

Binomial inverse problem:

Need to take the complement: i.e.

Cannot solve algebraically because we’re dealing with 30 discrete cases (0🡪29).

So need to approximate using the Normal Distribution, with

*Using a continuity correction.*

And since (from the info in the question), , then:

Need to find *n* such that  . i.e. from tables/GC, z= -1.2815

Therefore:

Squaring both sides.

And by the quadratic formula or GC polynomial solver, we get: ***n* = 80.743, n=119.75**

Given that *n* must be a positive whole number, round to ***n* = 81 or *n =* 120**. Try both of these values using the original Binomial Distribution problem:

**If *n* is 81**, we get: So

**If *n* is 120**, we get:

Therefore ***n* = 120** seems to work.

**\*\* Check on your GC using the Binomial Distribution:** Try **other** **integer *n* values near 120** because, remember, this solution was based on an *approximation*. Sure enough, 120 turns out to be the closest integer *n* value. But always pays to check. \*\*

**We need to sample at least 120 smokers in order to get a 0.9 probability that at least 30 give up.**

**Problem 11 (*Scholarship Statistics and Modelling 2010)***

On average, 3% of new Statsmobile cars have minor defects that occur randomly.

In a random sample of ***n*** new Statsmobiles, find ***n*** so that there is a 0.2 chance of no more than one Statsmobile having a minor defect.

Binomial distribution. Parameters are ***n*** : number of Statsmobiles in the sample (chosen randomly).

****** : probability that a randomly chosen Statsmobile has a defect.

where ********= 0.03**.

Let *X* : number of Statsmobiles in the sample that have defects.

We’re told that P(*X* < 1) = 0.2. Asked to find the required number of trials (Statsmobiles), ***n***.

P(*X* < 1) = P(*X*=0) + P(*X* = 1) where ********= 0.03**.

**METHOD A: Using Binomial Distribution directly, forming an equation and using technology to solve it.**

* ****P(*X* = 0) + P(*X* = 1) = 0.2

nC0 ×(0.03)0× (0.97)n-0 + nC1 ×(0.03)1× (0.97)n-1 = 0.2

0.97n + n × (0.03) × (0.97)n-1 = 0.2

Can solve using technology (Graphics Calculator) by going into GRAPH mode, then graphing 2 equations:

Y1 = 0.97***x*** + ***x*** × 0.03 × 0.97 ***x*** -1

Y2 = 0.2

Select **both** equations and, once graphed, click G-SOLVE, then ISCT to find the coordinates of the point of intersection.

The *x* value is the solution to the equation 0.97n + n × (0.03) × (0.97)n-1 = 0.2, and hence is the value of ***n*** that we’re after. This gives a solution of **98.808**. **And *n* must be a positive whole number. So best answer is *n* = 99 Statsmobiles in the sample.**

**METHOD B: Use a** ***Poisson Approximation***.

The reason why we must approximate to the Poisson rather than to the Normal distribution is because the probability ****** is very close to zero (0.03), and so the distribution will not be symmetrical but skewed to the right (peaked on the left close to zero).

Look up the Poisson probability tables, looking for *x* values of 0 and 1 where the probabilities sum to near 0.2.

Guess and check.

Closest is when ******= 3. Then, using the Poisson distribution, ****** **= n = 0.03n** *since =0.03*

**** **0.03n = 3**

**n = **

**= 100**

So, using this method & the Poisson tables, best answer **is *n* = 100** Statsmobiles in the sample.